



**You have downloaded a document from
RE-BUŚ
repository of the University of Silesia in Katowice**

Title: Zastosowanie struktur grafowych do analiz bibliometrycznych i webometrycznych. Modele i metody

Author: Anna Małgorzata Kamińska

Citation style: Kamińska Anna Małgorzata. (2018). Zastosowanie struktur grafowych do analiz bibliometrycznych i webometrycznych. Modele i metody. "Nowa Biblioteka" (Nr 2 (2018), s. 47-63)



Uznanie autorstwa - Użycie niekomercyjne - Bez utworów zależnych Polska - Licencja ta zezwala na rozpowszechnianie, przedstawianie i wykonywanie utworu jedynie w celach niekomercyjnych oraz pod warunkiem zachowania go w oryginalnej postaci (nie tworzenia utworów zależnych).



UNIwersYTET ŚLĄSKI
W KATOWICACH



Biblioteka
Uniwersytetu Śląskiego



Ministerstwo Nauki
i Szkolnictwa Wyższego



Anna Małgorzata Kamińska

Zakład Bibliotekoznawstwa
Instytut Bibliotekoznawstwa i Informacji Naukowej
Uniwersytet Śląski w Katowicach
e-mail: anna.kaminska@us.edu.pl

Zastosowanie struktur grafowych do analiz bibliometrycznych i webometrycznych Modele i metody

Abstrakt: W artykule przedstawiono możliwość wykorzystania koncepcji zaczerpniętych z teorii grafów oraz metod analizy sieci społecznościowych w zastosowaniach biblio- i webometrycznych. Chociaż metody analizowania struktur grafowych znane są w nauce od dłuższego czasu, to dopiero rozwój technologii pozwalających zastosować te metody w praktyce na dużych zbiorach danych spowodował ich upowszechnienie w różnych obszarach nauki, a nawet życia codziennego. Możliwości wykorzystania wspomnianych metod na gruncie biblio- i webometrii były sygnalizowane przez badaczy już w ubiegłym wieku, jednak w stosunkowo niewielu publikacjach, zwłaszcza krajowych, stosuje się te metody w praktyce. Dlatego w niniejszym opracowaniu omówione zostały ogólne koncepcje analiz i modeli agregacji danych bibliometrycznych, które mogą zostać zaadaptowane na potrzeby konkretnych badań. Rozważania te mają charakter generalny i niezależny od konkretnej platformy analitycznej, natomiast w końcowej części artykułu zaprezentowano przykład aplikacji Gephi, umożliwiającej implementację opisywanych koncepcji.

Słowa kluczowe: Bibliometria. Gephi. Grafy dynamiczne. Grafy hierarchiczne. Metody ewaluacji. Modele ewaluacji. Struktury grafowe

Wstęp

Bibliometria i webometria, jako dziedziny obejmujące swoim zakresem badania nad zależnościami występującymi pomiędzy dokumentami reprezentującymi wiedzę oraz nad zależnościami pomiędzy autorami

tych dokumentów, umożliwiają realizację wielu zadań. Wśród najważniejszych należy wymienić: ewaluację naukową autorów, czasopism, jednostek naukowych, badania nad rozwojem nauki i powstawaniem jej nowych gałęzi, wizualizację rozwojów dziedzin czy tworzenie map nauki. Więcej informacji na temat metod, zastosowań i interpretacji wyników bibliometrycznych i webometrycznych badań ilościowych znaleźć można np. w publikacji Ireny Marszakowej-Szajkiewicz (Marszakowa-Szajkiewicz, 2009), natomiast obszerny przegląd piśmiennictwa tych dziedzin został przedstawiony w artykule Łukasza Opalińskiego (Opaliński, 2017).

Niezależnie od tego, czy analizowanymi jednostkami będą wydawnictwa drukowane, publikacje elektroniczne, czy też strony internetowe, pryncypia są podobne, a polegają one na identyfikowaniu i analizowaniu związków pomiędzy poszczególnymi jednostkami, czyli jednostką badaną i tymi, które zostały przywołane w jej treści, stanowiąc inspirację dla autora, oraz związków pomiędzy autorami, którzy wnieśli wkład intelektualny w powstanie poszczególnych jednostek.

Metody analizowania wskazanych związków w naturalny sposób ewoluowały wraz z ogólnym rozwojem technologii informatycznych i metod obliczeniowych. Początkowo były to proste algorytmy operujące na zbiorze dwuargumentowych relacji jednostka – jednostka oraz jednostka – autor, a obecnie rozwinęły się do postaci złożonych metod obliczeniowych, których wyniki nie zależą jedynie od bezpośredniego sąsiedztwa pojedynczych obiektów (Wojciechowski, Pieńkosz, 2013). Należy bowiem zauważyć, że związki te, postrzegane całościowo, stanowią swoistą sieć połączeń, a dokładniej graf (najczęściej skierowany), którego wierzchołki reprezentują jednostki wiedzy, ich autorów bądź inne analizowane byty, zaś krawędzie – związki, które je łączą. Fakt ten dostrzegany był już od samego początku rozwoju metod bibliograficznych (Garfield, Sher, Torpie, 1964), jednak dopiero rozwój technologii i algorytmiki pozwolił wypracować metody wykorzystywane współcześnie.

Paradoksalnie implementacja takich struktur (umożliwiająca gromadzenie i składowanie danych) w konkretnych systemach zarządzania bazami danych była łatwiejsza i bardziej naturalna w przypadku mało popularnych sieciowych baz danych, których koncepcja pojawiła się jeszcze przed koncepcją baz relacyjnych (Navathe, 1992). Jednak to te ostatnie, ze względu na swoją uniwersalność i elastyczność, zyskały w ciągu kilkudziesięciu ostatnich lat największą popularność, co spowodowało, że platformy relacyjnych baz danych są obecnie najbardziej

dojrzałymi narzędziami służącymi do gromadzenia, przechowywania, przetwarzania i udostępniania danych.

Warto zauważyć, że wybrana dla danej bibliograficznej bazy danych technologia fizycznego gromadzenia i składowania informacji (np. czy jest to relacyjny czy sieciowy system zarządzania bazą danych) ma drugoplanowe znaczenie, pod warunkiem jednak, że logiczna architektura składowanych danych umożliwiać będzie trawersowanie krawędzi grafów, będących *de facto* ścieżkami cytowań (Kamińska, 2017c).

Znakomita większość bibliograficznych baz danych wykorzystuje relacyjne systemy zarządzania danymi, w związku z tym w celu realizacji wielu analiz bibliometrycznych zachodzi konieczność odwzorowania danych gromadzonych w układzie relacyjnym w dane reprezentujące struktury grafowe. W niniejszym opracowaniu przedstawiono sposoby tych odwzorowań, a główną motywacją autorki był dobór odwzorowań umożliwiających w miarę elastyczne prowadzenie badań, co w praktyce oznacza dobór jak najmniejszej liczby modeli o jak największym zakresie informacyjnym, ale jednocześnie nie większym, niż pozwoli na to stosowanie późniejszych miar bibliometrycznych (np. dla grafów wielomodalnych wyniki stosowania niektórych miar są trudno interpretowalne – zostanie to szczegółowo wyjaśnione w dalszej części opracowania).

Tematyka możliwości zastosowania elementów teorii grafów do prowadzenia badań nad rozwojem nauki była już sygnalizowana przez krajowych badaczy (Muraszkiewicz, Rybiński, Szczepański, 2015), którzy za pomocą języka teorii zbiorów przedstawili ważne grafy nieskierowane jako narzędzie wspierające analizy współtworzenia artykułów oraz współuczestnictwa w projektach. W niniejszym artykule autorka rozszerza te rozważania o nowe typy grafów, tzn. grafy skierowane, hierarchiczne i dynamiczne, pokazując na przykładach możliwości ich użycia w prowadzeniu analiz bibliometrycznych.

Drugie rozszerzenie wspomnianej pracy (Muraszkiewicz, Rybiński, Szczepański, 2015) stanowi propozycja metody opisu grafów. Autorka, zainspirowana powszechnym wykorzystywaniem notacji graficznych w projektowaniu relacyjnych struktur danych (zamiast pierwotnie używanego do tego celu matematycznego języka algebry relacji), postanowiła na gruncie modelowania struktur grafowych zrezygnować z języka matematycznego na rzecz graficznego języka UML (Unified Modeling Language). Takie obrazowe przedstawienie modeli może sprawić, że poruszana tematyka będzie przystępniejsza dla środowisk humanistycznych, a to z kolei może przyczynić się do wzrostu zainteresowania rozszerzeniem zakresu prowadzonych badań bibliometrycznych poza najczęściej stosowane układy tabelaryczne.

Trzecie rozszerzenie stanowi rozwinięcie zagadnień grafów wielomodalnych oraz wykazanie, że wiele miar do nich stosowanych, bez ich uprzedniej redukcji do grafów unimodalnych, może dawać wyniki trudne do interpretacji. Autorka na zilustrowanym graficznie przykładzie pokazuje sposób realizacji takiej redukcji.

Reprezentacja danych bibliograficznych za pomocą struktur grafowych

Na potrzeby niniejszych rozważań graf wystarczy zdefiniować jako zbiór wierzchołków, które mogą być połączone krawędziami w taki sposób, że każda krawędź zaczyna się i kończy w jakimś wierzchołku. Jeśli istotne jest, który z wierzchołków jest początkowy, a który końcowy, wówczas można mówić o grafie skierowanym. W przypadku zastosowań w dziedzinie bibliometrii wierzchołki grafu to najczęściej reprezentacje jakichś obiektów (np. artykułów, autorów), zaś krawędzie mogą obrazować interakcje między wierzchołkami (np. cytowania, autorstwa). Zarówno wierzchołki, jak i krawędzie w zależności od typu (reprezentowanej abstrakcji) mogą posiadać dodatkowe atrybuty – np. wierzchołek autora może mieć przypisany atrybut reprezentujący jego narodowość. Jeśli istnieje potrzeba wyrażenia siły interakcji, wówczas krawędź może być opisana liczbą reprezentującą jej siłę, np. liczbą cytowań danego artykułu przez inny artykuł, a graf taki nazywamy grafem ważonym.

Rozwinięcie powyższych, bardzo uproszczonych i nieformalnych definicji, z wykorzystaniem języka matematyki, można znaleźć w pracy Juliusza Kulikowskiego *Zarys teorii grafów. Zastosowania w technice* (Kulikowski, 1986).

Grafy hierarchiczne i dynamiczne

Wybrane narzędzia do analizowania struktur grafowych udostępniają funkcjonalność przetwarzania grafów dynamicznych i hierarchicznych. Grafy dynamiczne są osadzone w czasie, co w najprostszym przypadku oznacza, że wierzchołki lub krawędzie mają zdefiniowany czas lub okres życia. Bardziej zaawansowane narzędzia dopuszczają nawet zmienność w czasie pojedynczych atrybutów wierzchołków bądź krawędzi. Daje to możliwość zdefiniowania punktu w przestrzeni czasowej bądź zdefiniowania przedziału czasu, dla którego realizowane jest

przetwarzanie (analiza struktury). Pozwala to śledzić rozwój pewnych interakcji, identyfikować trendy czy odpowiadać na pytanie, w jakim czasie siła danej interakcji osiągała wartości ekstremalne.

Odrębną kategorię stanowią grafy hierarchiczne – w przypadku niektórych wierzchołków istnieje tu możliwość zdefiniowania wierzchołków nadrzędnych (tzw. rodziców). Pozwala to na budowanie struktur drzewiastych o zdefiniowanej liczbie poziomów, co sprawia, że analiza grafu może być zawężana do konkretnego poziomu albo do wszystkich wierzchołków niemających już wierzchołków podrzędnych (tzn. liści drzewa). Należy zwrócić uwagę, że krawędzie w takim grafie mogą być definiowane pomiędzy wierzchołkami tego samego poziomu (wtedy będą one widoczne i uwzględnione w przetwarzaniu grafu zawężonego do tego poziomu), mogą też być definiowane pomiędzy wierzchołkami różnych poziomów (wtedy nie będą widoczne ani uwzględnione w przetwarzaniu grafu zawężonego do któregośkolwiek z poziomów). Niektóre narzędzia udostępniają mechanizm tworzenia tzw. krawędzi wirtualnych, tzn. automatycznego tworzenia krawędzi na danym poziomie przetwarzania, wynikających z istnienia krawędzi na poziomach niższych. Mechanizm ten ilustruje rysunek 1.



Rys 1. Przykład mechanizmu tworzenia krawędzi wirtualnych

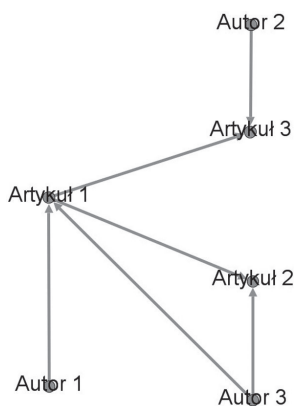
Źródło: Opracowanie własne.

Wierzchołki 1.1, 1.2 i 1.3 zostały zdefiniowane na poziomie 1 i połączone krawędziami o wartościach wag 1. Dla wierzchołków 1.1 i 1.2 zdefiniowano wierzchołek nadrzędny 0.12 (poziom 0), natomiast dla wierzchołka 1.3 zdefiniowano wierzchołek nadrzędny 0.3 (poziom 0). Pomiedzy wierzchołkami 0.12 i 0.3 nie zdefiniowano w sposób jawny żadnych wierzchołków, jednak stosując technikę krawędzi wirtualnych, otrzymamy krawędź o wadze 2, jako że siła oddziaływań pomiędzy grupą składającą się z wierzchołków 1.1 i 1.2 a jednoelementową grupą składającą się z wierzchołka 1.3 wynosi dokładnie 2. Krawędź łącząca wierzchołki 1.1 i 1.2 na poziomie 0 przestaje mieć znaczenie. Zastosowaną funkcją agregującą jest tutaj suma, jednak w zależności od typu oddziaływań można zastosować również średnią.

Nic nie stoi na przeszkodzie, aby traktować grafy hierarchiczne dodatkowo również jako dynamiczne.

Grafy wielomodalne i unimodalne

Możliwość zastosowania niektórych technik analitycznych uwarunkowana jest odpowiednią budową struktur danych (grafów). Często zdarza się, że wierzchołki grafu reprezentują różne klasy bytów (można wtedy mówić o grafach wielomodalnych). W przypadku analizowania sieci cytowań mogłyby to być przykładowo wierzchołki reprezentujące artykuł bądź autora. Natomiast krawędzie wskazywałyby cytowanie bądź autorstwo. Przykładowy graf przedstawiono na rysunku 2.



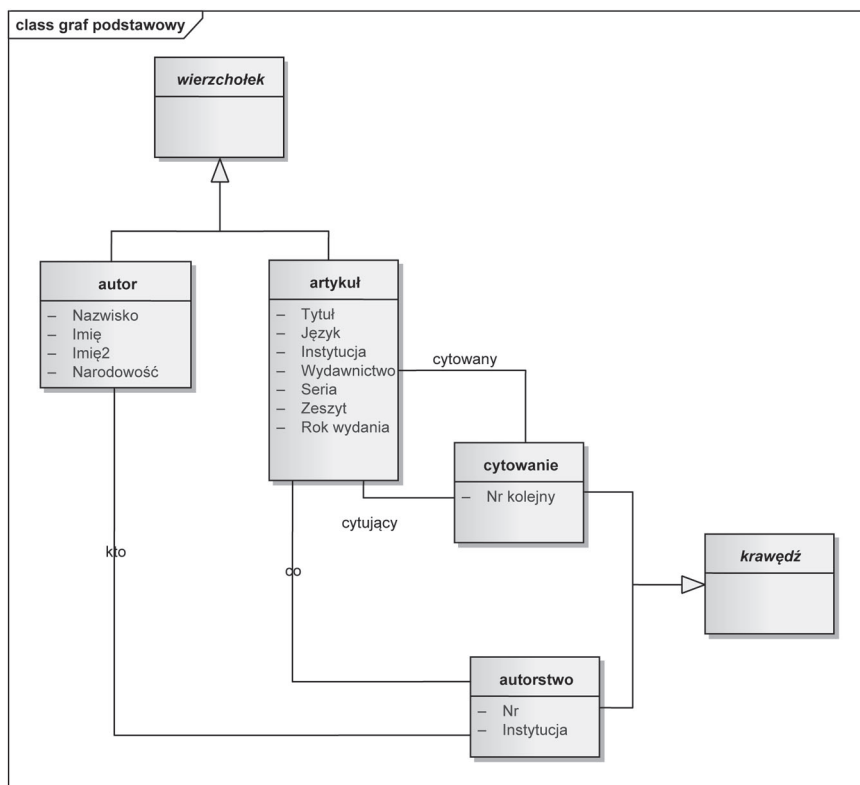
Rys. 2. Przykładowy graf autorstwa i cytowań

Źródło: Opracowanie własne.

Jest to graf skierowany, z którego wynika, że w Artykule 1 cytowany jest Artykuł 2 i Artykuł 3. Autor 2 jest autorem Artykułu 3. Autorami Artykułu 1 są Autor 1 i Autor 3 (jest on również autorem Artykułu 2). Ten szczególny przypadek można uogólnić w postaci diagramu klas przedstawionego na rysunku 3.

Więcej informacji o metodzie modelowania za pomocą diagramów klas języka UML na przykładzie dziedziny bibliologii można znaleźć w artykule o graficznych językach komunikacji (Kamińska, 2018a) natomiast obszernie omówienie języka UML – w opracowaniu jego twórców (Booch, Rumbaugh, Jacobson, 2002).

Diagram klas definiujący budowę grafu należy interpretować w ten sposób, że wierzchołki grafu mogą reprezentować autora (wtedy dodatkowymi atrybutami opisującymi go mogą być: nazwisko, imiona, narodowość itd.) albo artykuł (wtedy dodatkowymi atrybutami opisującymi go mogą być: tytuł, język, rok wydania itd.), natomiast krawędzie mogą reprezentować cytowanie lub autorstwo. Krawędzie oznaczające cyto-



Rys. 3. Przykład diagramu klas definiującego budowę grafu

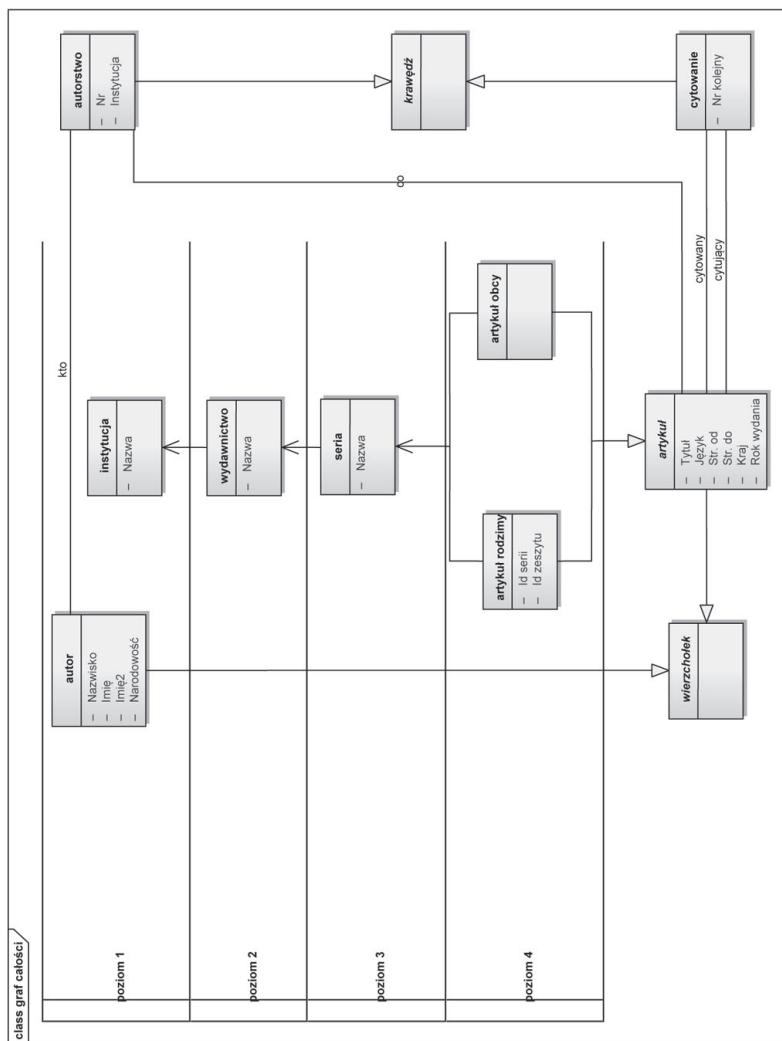
Źródło: Opracowanie własne.

wanie łączą dwa wierzchołki typu artykuł, gdzie wyróżnia się artykuł cytowany i cytujący (a więc jest to graf skierowany). Krawędzie oznaczające autorstwo łączą wierzchołki typu autor z wierzchołkiem typu artykuł i wskazują, kto jest autorem czego (podobnie jak w przypadku poprzedniego typu krawędzi, tutaj też istotny jest jej kierunek). Przykładowe atrybuty umieszczone w klasach krawędzi mogą nieść dodatkowe informacje rejestrowane w trakcie gromadzenia danych. I tak np. krawędź oznaczająca cytowanie zawiera atrybut 'Nr kolejny' – numer wystąpienia cytowanego artykułu w bibliografii załącznikowej artykułu cytującego. Niektóre z tych informacji mogą być pomocne podczas prac analitycznych, większość narzędzi do analizy struktur grafowych umożliwia bowiem realizację takich operacji, jak np. filtrowanie czy kolorowanie grafu według przyjętych kryteriów zdefiniowanych na takich atrybutach.

Choć badaczom zajmującym się bibliometrią przykład przedstawiony na rysunku 2 może wydać się trywialny, to bezpośrednie stosowanie większości miar ilościowych do przedstawionego na nim grafu może dawać trudno interpretowalne wyniki, gdyż mierzone byłyby wspólnie różne rodzaje oddziaływań między różnymi rodzajami bytów. Na przykład mierząc stopień wierzchołka dla Artykułu 1, otrzymamy wartość 4, w której zawierać się będzie zarówno liczba cytowań, jak i 'autorstw'. Dlatego ze względu na różnorodność celów różnych analiz istnieje potrzeba konstruowania dedykowanych struktur grafowych – przykładowo chcąc obserwować, jak kształtowała się współpraca między poszczególnymi autorami, należałoby skonstruować graf o wierzchołkach typu autor, a krawędzie powinny w nim oznaczać fakt współpracy (współautorstwa, ang. *co-authorship*). Natomiast chcąc analizować rozwój dziedzin nauki, należałoby stworzyć graf z wierzchołkami reprezentującymi artykuły, w którym krawędzie oznaczałyby cytowania wynikające z bibliografii załącznikowej. Krótko mówiąc, najpierw należałoby przeprowadzać transformacje grafu wielomodalnego do grafów unimodalnych, a następnie na tak otrzymanych grafach dokonywać odpowiednich analiz.

Nie oznacza to jednak, że konstruowanie grafów zawierających wierzchołki i krawędzie wielorakich typów zawsze jest pozbawione sensu. W zasadzie tylko taki graf, o strukturze maksymalnie zbliżonej do modelu, na podstawie którego dane były gromadzone, daje możliwość przeglądu wszystkich informacji w ramach jednej struktury (gdyż przekształcając grafy wielomodalne w unimodalne, w oczywisty sposób tracimy pewne informacje). Co więcej, struktura ta często może być zredukowana do postaci, o których wspomniano wcześniej, za pomocą mechanizmów filtrowania i redukowania grafów, udostępnianych przez większość narzędzi do ich analizowania. Pełniejsza przykładowa struktura, zaimplementowana dla własnych badań bibliometrycznych przeprowadzonych przez autorkę, została przedstawiona na rysunku 4.

Grafy skonstruowane według przykładowego modelu (obejmującego zakresem informacyjnym dziedzinę wydawnictw ciągłych), zawierające całość gromadzonych informacji, można łatwo zredukować do grafów, które mogą stanowić podzbiory analizowane za pomocą różnych narzędzi. Jak widać, krawędzie reprezentujące cytowania definiowane są zawsze w odniesieniu do artykułów, czyli na czwartym poziomie hierarchii. Zawężając graf do tego poziomu, otrzymamy więc graf cytowań artykułów.



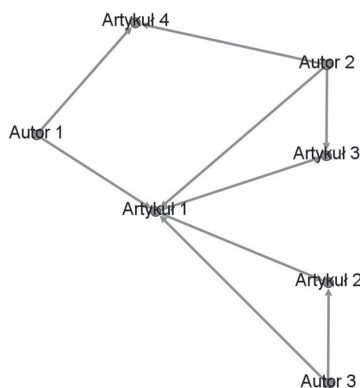
Rys. 4. Model grafu o całościowym zakresie informacyjnym

Źródło: Opracowanie własne.

Zawężając graf do poziomu trzeciego i drugiego, z wykorzystaniem techniki krawędzi wirtualnych, można analizować cytowania odpowiednio między seriami i wydawnictwami. Na poziomie pierwszym możemy obserwować autorów i instytucje, a filtrując graf tylko do instytucji, możemy obserwować cytowania między instytucjami. Należy zwrócić uwagę, że w tak zdefiniowanym grafie hierarchicznym występują jedynie dwa rodzaje liści (elementy struktur drzewiastych nieposiadające dalszych gałęzi) – są to artykuły i autorzy. Filtrując graf tak, by przedstawiał tylko liście hierarchii, otrzymamy widok artykułów i autorów z krawędziami cytowań i ‘autorstw’ (dzięki krawędzinom wirtualnym przebiegającym z poziomu 4 do 1). Wystarczy teraz odrzucić krawędzie cytowań, aby otrzymać unimodalny graf przedstawiający autorstwa artykułów.

Jak widać, opisana struktura jest wystarczająco elastyczna, aby umożliwić analizowanie zarówno cytowań artykułów, jak i agregacji na poziomach wyższych, takich jak serie, wydawnictwa czy instytucje. Umożliwia również analizowanie wkładu autorskiego konkretnych osób: kto tworzył/współtworzył artykuły, ile artykułów powstało, w tworzenie których artykułów zaangażowana była największa liczba osób itd. Dzięki atrybutowi określone jako rok w wierzchołkach artykułów istnieje możliwość realizacji analiz przekrojowych dotyczących wybranych lat bądź też inaczej zdefiniowanych okien czasowych.

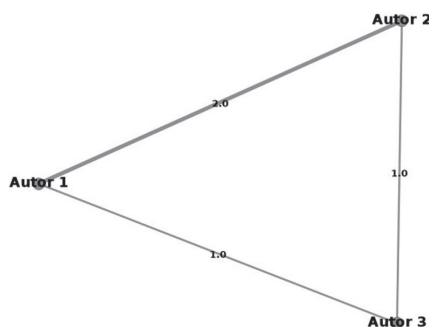
Mimo że opisana jako ostatnia redukcja grafu umożliwia obserwowanie, kto jest autorem jakiego artykułu, i mimo że zawarta w niej jest informacja, którzy autorzy ze sobą współpracowali, to w dalszym ciągu jest to graf o dwóch typach wierzchołków i nie można do niego stosować miar bezpośrednich. Transformacja takiego grafu jest prosta, a przykład został przedstawiony na rysunku 5.



Rys. 5. Przykład grafu bimodalnego przed transformacją

Źródło: Opracowanie własne.

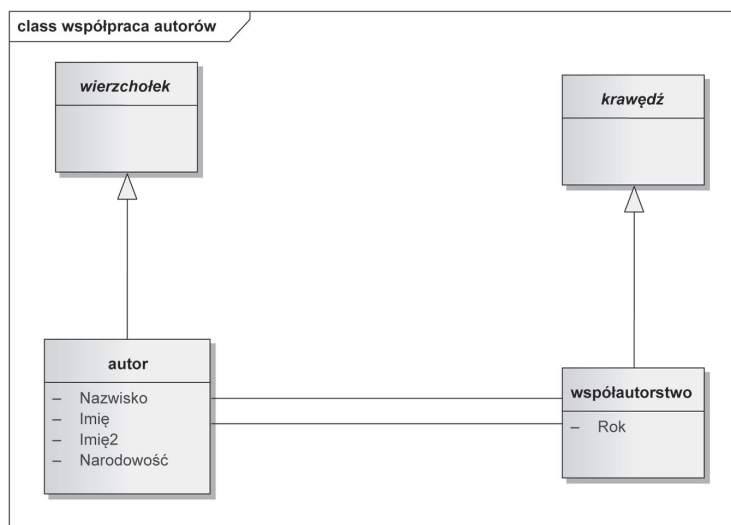
Autor 1 i Autor 2 współpracowali przy Artykule 4 i Artykule 1. Natomiast Autor 3 i Autor 1 współtworzyli Artykuł 1. Artykuł 2 i Artykuł 3 nie wnoszą nic do grafu współpracy, gdyż posiadają tylko jednego autora. Tak więc po transformacji otrzymujemy graf zaprezentowany na rysunku 6.



Rys. 6. Przykład grafu unimodalnego po transformacji

Źródło: Opracowanie własne.

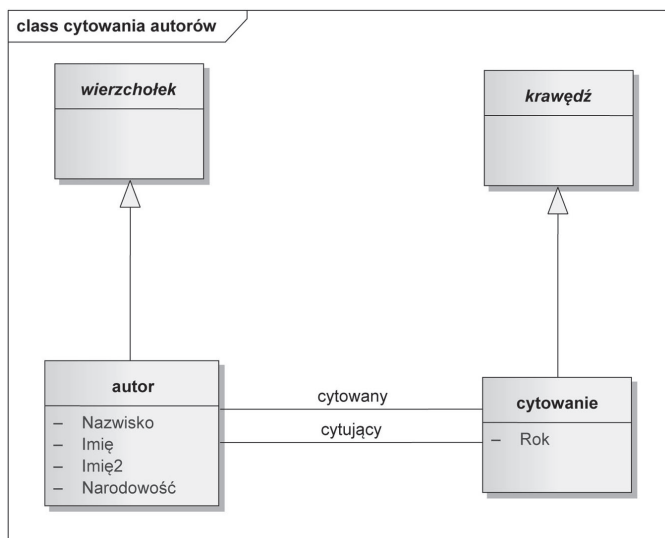
Należy zwrócić uwagę, że z nieważonego skierowanego grafu powstał graf ważony (gdzie waga pokazuje siłę współpracy – w tym przypadku liczbę wspólnie tworzonych artykułów) i nieskierowany, gdyż relacja współautorstwa jest relacją zwrotną. Strukturę takiego grafu, zdefiniowaną za pomocą diagramu klas, przedstawiono na rysunku 7.



Rys. 7. Model grafu współpracy autorów

Źródło: Opracowanie własne.

Z podobną sytuacją mamy do czynienia w przypadku redukcji grafu całości gromadzonych informacji do grafu cytowań między autorami. Jedyna różnica polega na tym, że relacja cytowania nie jest zwrotna, więc powstająca struktura jest grafem skierowanym (nazwane relacje między krawędziami a wierzchołkami). Model takiego grafu przedstawiono na rysunku 8.



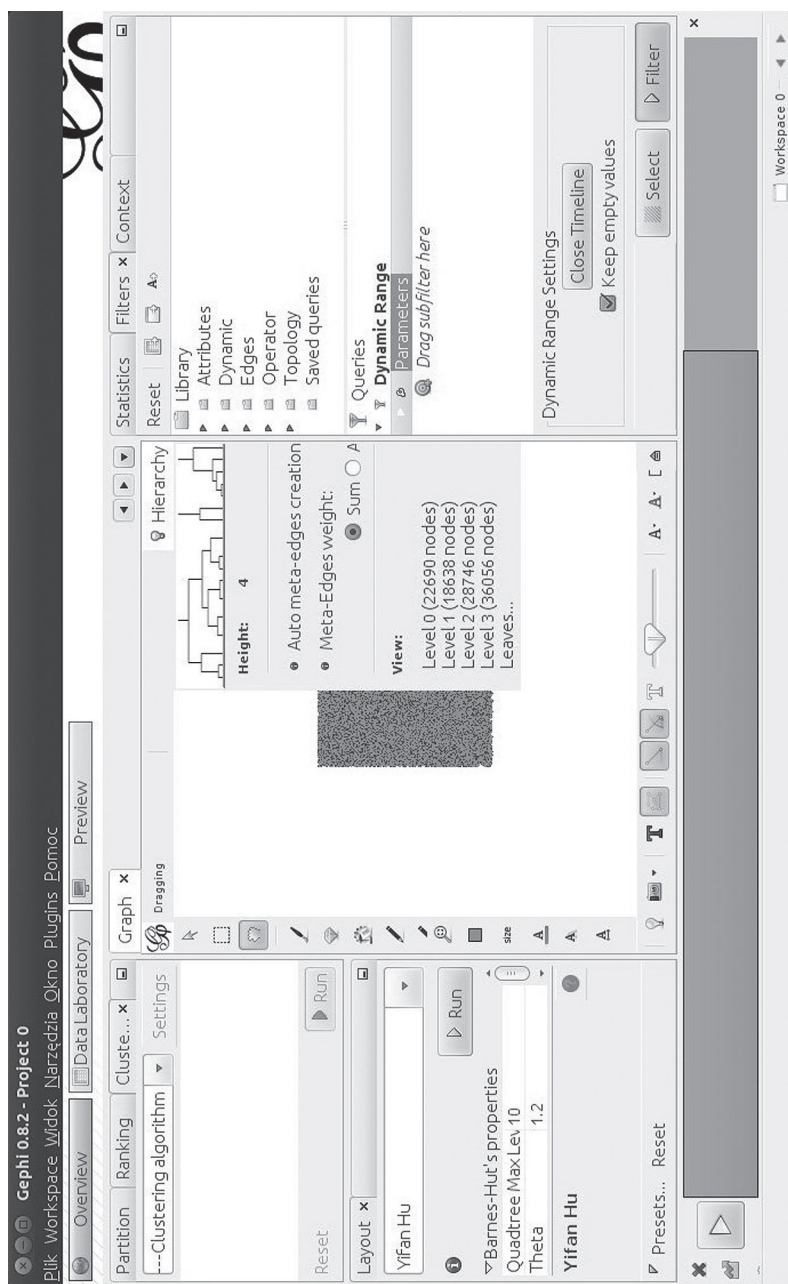
Rys. 8. Model grafu cytowań autorów

Źródło: Opracowanie własne.

Zdefiniowane modele grafów stanowiły podstawę bardziej zaawansowanych analiz bibliometrycznych realizowanych w ramach badań autorskich, które zostaną przedstawione w kolejnych publikacjach. Zakres informacyjny poszczególnych typów wierzchołków i krawędzi należy traktować poglądowo i może się on zmieniać w zależności od potrzeb konkretnych analiz.

Przykład platformy do analizowania struktur grafowych

Na rysunku 9 przedstawiono główne okno programu analitycznego Gephi v0.8.2, który był wykorzystywany przez autorkę do implementacji koncepcji omówionych w niniejszym artykule. Po odpowiedniej transformacji danych bibliograficznych zgromadzonych w relacyjnej bazie danych zostały zaprojektowane struktury grafowe w układach



Rys. 9. Główne okno programu analitycznego Gephi (wersja 0.8.2)

Źródło: Opracowanie własne.

hierarchicznych i dynamicznych. Po wyeksportowaniu danych do takich układów i wczytaniu ich w programie Gephi system automatycznie rozpoznaje, że są to dane zorganizowane właśnie w ten sposób, i udostępnia możliwość ich hierarchicznej agregacji i zawężania (filtrowania) według zdefiniowanych okien czasowych. W środkowym górnym miejscu okna znajduje się dialog 'Hierarchy' informujący o liczbie poziomów hierarchii (w tym przypadku czterech) oraz o liczbie wierzchołków na każdym z poziomów. Agregację przeprowadza się, klikając bezpośrednio wybrany poziom lub opcję 'Leaves' przełączającą graf w tryb widoku liści. Jak widać, istnieje tutaj również możliwość włączenia ustawienia tworzenia krawędzi wirtualnych oraz określenia używanej miary agregującej (Sum/Average).

Podsumowanie

W artykule przedstawiono możliwości analiz bibliometrycznych wykorzystujących koncepcje zaczerpnięte z teorii grafów. Gromadzenie danych bibliograficznych w relacyjnych bazach danych (a takich właśnie platform w znakomitej większości używają bibliograficzne bazy danych) w naturalny sposób narzuca przetwarzanie tych danych w układach tabelarycznych (relacyjnych). Może to skutkować sprowadzeniem analiz ilościowych jedynie do prostych podsum, sortowań czy agregacji, co z kolei ogranicza możliwości szerszego wnioskowania i obserwacji zjawisk zachodzących w rozwoju nauki. Autorka dzieli się uwagami na temat postrzegania danych bibliograficznych (nawet tych gromadzonych w układach relacyjnych) jako struktur grafowych, co umożliwia szersze spojrzenie na zjawiska zachodzące w społecznościach naukowych i otwiera nowe możliwości ich analiz. Pokazuje również, że dzięki wykorzystaniu koncepcji grafów dynamicznych i hierarchicznych analizowanie danych w układach grafowych nie wyklucza możliwości stosowania operacji agregacji i selekcji (zawężania do przedziałów czasowych lub wierzchołków o zadanych atrybutach), analogicznych do tych wykorzystywanych przy analizach w tradycyjnych układach relacyjnych.

Przedstawiona koncepcja, jak również konkretne propozycje budowania struktur grafowych umożliwiają realizację badań bibliometrycznych o wybranych profilach, np. badania współpracy między autorami (współautorstwa), cytowań między autorami czy cytowań między poszczególnymi artykułami naukowymi. Budując takie struktury, należy pamiętać o możliwości wystąpienia „efektów ubocznych” w przypadku

analizowania grafów wielomodalnych – w postaci możliwości uzyskania trudno interpretowalnych wyników. W takich przypadkach należy zastosować techniki redukcji grafów wielomodalnych do postaci unimodalnych, umożliwiających realizację badań w wybranych perspektywach (przekrojach). Zostało to zilustrowane na konkretnych przykładach uwzględniających autorów, artykuły, autorstwa i cytowania.

Zaprezentowany przykład narzędzia analitycznego może zostać wykorzystany do realizacji analiz bibliometrycznych z bezpośrednim zastosowaniem przedstawionych koncepcji. Praktyczna przydatność tych koncepcji oraz samego narzędzia została zweryfikowana podczas badań autorskich. Obejmowały one analizy cytowań na poziomie artykułów dla czasopisma PLOS ONE na podstawie danych zaczerpniętych z indeksu cytowań OpenCitations Corpus, wykorzystującego grafową bazę danych (Kamińska, 2017a), na poziomie cytowań pomiędzy autorami i pomiędzy czasopismami na podstawie danych wyekstrahowanych z krajowej bibliograficznej bazy danych CYTBIN (Kamińska, 2017b; Kamińska, 2017d) oraz analizy relacji współautorstwa na podstawie samodzielnie opracowanej bibliograficznej bazy danych gromadzącej informacje z dziedziny górnictwa rozwijającego się w Polsce w latach 1945–1989 (Kamińska, 2018b).

Treści zawarte w niniejszym opracowaniu mogą być wykorzystane na różnych etapach realizacji badań bibliometrycznych poprzez:

- uwzględnienie sieciowej natury danych bibliograficznych w trakcie opracowywania modelu gromadzonych danych;
- dostrzeganie sieciowych związków między zgromadzonymi danymi w celu sprofilowania kierunków realizowanych badań bibliometrycznych;
- opracowywanie modeli sieciowych na podstawie danych relacyjnych gromadzonych w konkretnych układach, a następnie eksportowanie ich do dedykowanych narzędzi wspierających realizację badań nad strukturami sieciowymi;
- wykorzystanie narzędzi wspierających realizację ww. badań w celu transformacji grafów (np. unimodalnych) do innych postaci, bardziej przyjaznych w realizacji konkretnego profilu badań.

Dalsze prace autorki w tej dziedzinie w ramach poruszanych zagadnień koncentrować się będą na badaniach możliwości analitycznych dostępnych narzędzi w zakresie ich zastosowania w bibliometrii oraz na możliwościach interpretacji na gruncie tej dziedziny wyników otrzymanych z wykorzystaniem metod ogólnych, tzn. przeznaczonych do analizowania struktur sieciowych jako takich, bez szczególnego uwzględnienia dziedziny przedmiotowej przez nie reprezentowanej.

Literatura

- Booch, G., Rumbaugh, J., Jacobson, I. (2002). *UML przewodnik użytkownika*. Warszawa: Wydawnictwo Naukowo-Techniczne.
- Garfield, E., Sher, I., Torpie, R. (1964). *The use of citation data in writing the history of science. A final report of research for the Air Force Office of Scientific Research under contract AF 49(638)-1256*. Philadelphia: Institute for Scientific Information, Inc.
- Kamińska, A.M. (2017a). PLOS ONE – studium przypadku analizy cytowań prac naukowych na podstawie danych otwartego indeksu cytowań (Open-Citations Corpus). *Biuletyn EBIB*, (6). Pobrane z: <http://open.ebib.pl/ojs/index.php/ebib/article/view/564>.
- Kamińska, A.M. (2017b). Potencjał bibliometryczny bibliograficznej bazy danych CYTBIN w świetle prostych i złożonych wskaźników analitycznych. *Bibliotheca Nostra*, (2), 83–94.
- Kamińska, A.M. (2017c). ProBIT – prospektywna metoda tworzenia traversowalnych indeksów cytowań a współczesne problemy organizacji przestrzeni informacji w tradycyjnych bibliograficznych bazach danych. *Zagadnienia Informacji Naukowej*, 55(1), 66–82.
- Kamińska, A.M. (2017d). Wizualizacje wybranych wskaźników bibliometrycznych na przykładzie bibliograficznej bazy danych CYTBIN. *Toruńskie Studia Bibliologiczne*, 10(2), 163–187.
- Kamińska, A.M. (2018a). O rozwoju graficznych języków komunikacji. Przykład wykorzystania UML w obszarze dziedzin humanistycznych. *Zagadnienia Informacji Naukowej*, (2).
- Kamińska, A.M. (2018b). Visualizations of the GRUBA bibliographic database: From printed sources to the maps of science. In: V. Osinska, G. Osinski (eds.), *Information Visualization Techniques in the Social Sciences and Humanities*. Hershey, PA: IGI Global, 151–174. DOI: 10.4018/978-1-5225-4990-1.ch009.
- Kulikowski, J.L. (1986). *Zarys teorii grafów. Zastosowania w technice*. Warszawa: Państwowe Wydawnictwo Naukowe.
- Marszakowa-Szajkiewicz, I. (2009). *Badania ilościowe nauki. Podejście bibliometryczne i webometryczne*. Poznań: Uniwersytet im. Adama Mickiewicza.
- Muraszkiewicz, M., Rybiński, H., Szczepański, P. (2015). Discovering Research Collaboration Networks from Scientific Digital Libraries. *Zagadnienia Informacji Naukowej*, 53(2), 7–17.
- Navathe, S.B. (1992). Evolution of Data Modeling for Databases. *Communications of the ACM*, 35(9), 112–123.

- Opaliński, Ł. (2017). Bibliometryczna metodologia prognozowania i oceny rozwoju dyscyplin naukowych. Analiza piśmiennictwa. Cz. 1. Publikacje pionierskie, metoda powiązań bibliograficznych, metoda współcytowań i metoda współwystępowania specjalistycznej terminologii naukowej. *Zagadnienia Informacji Naukowej*, 55(1), 34–65.
- Wojciechowski, J., Pieńkosz, K. (2013). *Grafy i sieci*. Warszawa: Wydawnictwo Naukowe PWN.

Tekst w wersji poprawionej wpłynął do redakcji 23 stycznia 2018 r.

Anna Małgorzata Kamińska

Department of Library Science

Institute of Library Science and Scholarly Information

University of Silesia in Katowice

e-mail: anna.kaminska@us.edu.pl

The application of graph structures to the bibliometrical and webometrical analyses Models and methods

Abstract: The article presents the possibility to use the concept derived from the theory of graphs and methods of the social network analysis in bibliometrics and webometrics. Although the methods of the analysis of graph structures have been familiar to scholarship for a long time, it was the development of technologies which enable the application of these methods in practice in large collections of data that popularised them in many fields of science, and even of everyday life. Although the possibilities of the application of these methods in biblio- and webometrics were indicated by researchers even in the past century, there is a relatively small number of publications, especially released in this country, which put these methods to practice. Therefore in the present work the author presents the general concepts of the analyses and the models of aggregation of bibliometrical data, which may be adopted for the purposes of specific research. Although the considerations which are described are general and are independent of a specific analytical platform, the final part of the article presents an example of the Gephi application, which enables the implementation of the concepts in question.

Keywords: Bibliometrics. Dynamic graphs. Gephi. Graph structures. Hierarchical graphs. Methods of evaluation. Models of evaluation